

Article

Modeling Deep Neural Networks to Learn Maintenance and Repair Costs of Educational Facilities

Jimyong Kim ¹, Sangguk Yum ², Seunghyun Son ³, Kiyoung Son ⁴ and Junseo Bae ^{5,*}

¹ Department of Architectural Engineering, Mokpo National University, Mokpo 58554, Korea; jimy6180@gmail.com

² Department of Civil Engineering, Gangneung-Wonju National University, Gangneung 25457, Korea; skyeom0401@gwnu.ac.kr

³ Department of Architectural Engineering, Kyung Hee University, Suwon 17104, Korea; seunghyun@khu.ac.kr

⁴ School of Architectural Engineering, University of Ulsan, Ulsan 44610, Korea; sky9852111@ulsan.ac.kr

⁵ School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley PA1 2BE, UK

* Correspondence: junseo.bae@uws.ac.uk

Abstract: Educational facilities hold a higher degree of uncertainty in predicting maintenance and repair costs than other types of facilities. Moreover, achieving accurate and reliable maintenance and repair costs is essential, yet very little is known about a holistic approach to learning them by incorporating multi-contextual factors that affect maintenance and repair costs. This study fills this knowledge gap by modeling and validating deep neural networks to efficiently and accurately learn maintenance and repair costs, drawing on 1213 high-confidence data points. The developed model learns and generalizes claim payout records on the maintenance and repair costs from sets of facility asset information, geographic profiles, natural hazard records, and other causes of financial losses. The robustness of the developed model was tested and validated by measuring the root mean square error and mean absolute error values. This study attempted to propose an analytical modeling framework that can accurately learn various factors, significantly affecting the maintenance and repair costs of educational facilities. The proposed approach can contribute to the existing body of knowledge, serving as a reference for the facilities management of other functional types of facilities.

Keywords: educational facilities; deep learning; deep neural network; maintenance and repair cost; facilities management



Citation: Kim, J.; Yum, S.; Son, S.; Son, K.; Bae, J. Modeling Deep Neural Networks to Learn Maintenance and Repair Costs of Educational Facilities. *Buildings* **2021**, *11*, 165. <https://doi.org/10.3390/buildings11040165>

Academic Editor: Carlos Oliveira Cruz

Received: 28 March 2021

Accepted: 14 April 2021

Published: 15 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Facility maintenance and repair management aim to appropriately operate and maintain the function of a building by responding to the requirements of the users, correcting design and construction errors in the facility, and proactively preventing the aging of the building from ineffective operation and maintenance management [1]. More effective and efficient facility maintenance and repair management strategies have been underlined because the function and size of buildings have become more advanced and larger with their longer lifespans. In detail, the maintenance and repair phase accounts for about 85% of the life cycle cost of a facility asset, representing the biggest economic impact throughout its whole life cycle [1]. Nevertheless, it is difficult to maintain and improve upon the initial performance of a facility. A large gap between the expected resource input and the actual resource input (e.g., time and cost) often exists because a significant amount of data generated during the maintenance and repair phase of the actual facility are not used appropriately [2]. In addition, the relevant regulations and management manuals differ from each institution or facilities management company. In turn, conventional budget plans have often been developed and implemented, mainly reflecting the client's subjective judgment and wishes [3]. More recently, even though emerging technologies associated with the

fourth industrial revolution (e.g., the internet of things, artificial intelligence, immersive technologies such as augmented/virtual reality, smart construction machinery) have been widely and rapidly applied for the phase of design and construction, these technologies are still slowly implemented into the maintenance and repair phase of existing facilities. Furthermore, most previous studies on facility maintenance and repair management systems mainly dealt with facilities, the aging of facilities, safety, energy, and the safety of the facility users. For comprehensive facility maintenance and repair systems, it is necessary to consider not only various everyday safety concerns (e.g., vandalism, fire, crime, power outages, etc.) but also natural hazards that are often overlooked because their frequency is low—though the damage severity is high. Natural hazard management should also be conducted to identify, prepare for, and reduce the potential risks associated with natural hazards (e.g., floods, typhoons, hail, earthquakes, heavy snow, lightning, etc.) [4]. In this sense, in order to develop a proper mid- and long-term management plan underpinned by predictive maintenance approaches, it is important to systematically incorporate those multi-contextual characteristics of facilities into the maintenance and repair cost prediction modeling process.

In the context of educational facilities, it is essential to secure and maintain the functions of appropriate facilities. Thus, the quality of education is closely tied with the quality of educational facilities, and investment and maintenance in such facilities should be achieved in order to maintain an appropriately high quality of education [4]. However, although enormous amounts of money are spent on facility maintenance every year, it is difficult to maintain the original functions of educational facilities in accordance with the needs of the trainees with existing post-maintenance systems [2]. Therefore, it is essential to prepare for damage to facilities in advance to plan for all possible types of maintenance needs, and to hold budgetary reserves to address such damages. For this reason, the establishment of advanced strategies can be key to the successful budget management of educational facilities. In response to these needs, the industry is concentrating on building integrated facility maintenance and repair management systems to safeguard educational assets. However, there are many challenges associated with estimating costs for facility maintenance and repair [5]. One key reason is that facility maintenance and repair management include a wide range of expenditures for the maintenance and repair of the facility [6]. For example, accidents, fires, vehicular accidents, vandalism, and crimes may all affect the use of the buildings, while leaks may occur due to the deterioration or failure of the facilities and equipment, the overflowing of toilets and drainage pipes, power failure, sprinkler failure, or structural damage. Natural disasters also result in the need for maintenance work, such as hurricanes, heavy snowfall, rainfall, hail, lightning, and so on [7]. In addition, refurbishing or remodeling the facilities is sometimes preferred over maintenance and repair, and thus, it is difficult to make a better-informed decision on more feasible budgeting [8].

Since educational facilities have a longer life cycle than most other types of facilities, there remains an especially high degree of uncertainty in predicting maintenance and repair costs [2]. In addition, educational institutions seem to have relatively little interest in the maintenance of their facilities compared to the managers of other residential or commercial facilities. Although the maintenance and repair of educational facilities greatly affect the satisfaction and sense of achievement of the educators, students, and faculty, 25% of educational facilities are not suitably maintained or repaired and, on average, these buildings age at a faster rate than other types of facilities [9].

2. Point of Departure: Gaps in Existing Knowledge

The need for proper educational facility maintenance and repair strategies that can enhance the effectiveness of education and decrease overall maintenance-related expenditures is gradually increasing. However, many students have still led their lives in poorly managed educational facilities [9]. To address this, many research efforts were made. The existing research on facility operations and maintenance has focused on performance

evaluation. For example, Baldry et al. (2000) developed a framework based on a balanced scorecard and suggested a method of assessing facility performance that is related to the characteristics of education [6]. Kaplan and Norton (2005) proposed the establishment of short-term and long-term operation strategies for educational facilities from the perspectives of users, management, learning, and finance [10]. Kok et al. (2011) investigated the function of facility operations and maintenance for an educational environment through various literature reviews, and analyzed the relationship between the quality of education and performance of maintenance [11]. Tamosaitiene et al. (2013) identified various service factors, such as the security level in a building, cleaning of the territory and building, general management, and maintenance periods, and developed a maintenance assessment program, applying game theory [12]. Lavy et al. (2014) identified functions, users, and financial factors to evaluate the performance of facility operations and maintenance [13]. Other studies were also focused mainly on evaluating the potential risks within certain facilities and their impacts on maintenance costs associated with a limited number of factors, such as safety, accidents, and fires [1,6,14–18].

For effective facility operations and maintenance management, not only facility replacement/repair and equipment deterioration but also a variety of factors that can decrease the functioning of the actual facility (e.g., accidents, fires, crimes, theft, vandalism, typhoons, earthquakes, floods, hail, heavy snow, and other natural disasters) should be planned for in advance. However, very little is still known about the impacts of these critical factors affecting maintenance and repair, although these potential risks are likely to lead to the deterioration or malfunction of some part of a facility or of certain functions. To prevent this outcome, potential risk factors should be identified in advance, and the possible extent of financial damage should be reduced in advance.

In addition, to estimate the facility maintenance and repair costs associated with these various potential risk factors, data on the potential costs during the whole life cycle of the facility and the proper tools that are capable of quantitatively analyzing such costs are required. However, to the best knowledge of the research team, existing facility operation management research is far from empirical, as most of the calculations of each potential risk factor are conducted based on literature surveys, questionnaires, and experts' advice [1,18,19]. In turn, very little is still known about analytical modeling frameworks that are underpinned by multiple factors affecting facility maintenance and repair costs.

Recently, technological advancements, such as big data, robotics, artificial intelligence, and unmanned transportation, have been introduced and applied in a wide range of fields such as finance, medical care, education, and construction, and their effects have been widely recognized [20,21]. To fill the gaps in the existing knowledge delineated above, applying one of these emerging technologies for developing a robust modeling framework is needed.

3. Research Objectives and Method

In order to reduce any potential financial losses during the operation and maintenance phase, it is essential to reliably and accurately predict the maintenance and repair costs of educational facilities. To this end, the main objective of this study was to develop and validate a deep neural network model to efficiently and accurately learn the maintenance and repair costs of educational facilities, drawing on 1213 high-confidence data points. The developed model learns and generalizes claim payout records on the maintenance and repair costs from facility asset information, geographic profiles, natural hazard records, and other causes of financial losses. In this paper, we propose a new modeling approach that can accurately learn various factors that significantly affect the maintenance and repair costs of educational facilities. The objectives of this study were achieved based on the following five steps:

- (1) As the claim payouts were processed according to standardized procedures, objective damage analysis could be performed by a certified loss adjuster to calculate damages, which is highly reliable [22]. In this sense, insurance claim payouts associated with

maintenance and repair records in educational facilities were collected from the Canadian University Mutual Insurance Exchange (CURIE) 1988–2018.

- (2) Maintenance and repair track records, basic information (building area, number of students, geographic locations), facility asset information (capital assets, building cost, building age), and information on the natural hazards in surrounding areas of the educational facilities were then collected.
- (3) A deep neural network model was developed by evaluating the learning performance of 14 different network alternatives that learn the maintenance and repair costs of educational facilities from multi-contextual factors gathered in Step 2. In this study, Python 3.7 was used with the Keras and Scikit-Learn libraries.
- (4) For the purpose of validation, a multiple regression analysis was conducted separately, using the IBM Statistical Package for the Social Sciences (SPSS) V22.
- (5) The robustness of the developed learning model was then validated by comparing it with the multiple regression analysis results based on their root mean square error and mean absolute error values.

4. Data Collection and Classification

This study collected claim payout records from 1988 to 2018 from the Canadian University Mutual Insurance Exchange (CURIE). CURIE has been a Canadian university insurance provider since 1988. In addition to providing general liability insurance, the organization compensates for physical losses, such as loss of equipment, property, and automobiles. In this study, the maintenance and repair costs per educational facility were extracted from the claim payout amounts of the maintenance and repair records and defined as the amount ratio to learn, seen in Equation (1).

$$\text{Maintenance and repair cost ratio} = \frac{\text{Maintenance and repair cost (\$)}}{\text{Total cost of educational facility (\$)}} \quad (1)$$

Based on the collected claim payout records, maintenance and repair history and basic information regarding the educational facilities (i.e., capital assets, number of students, geographic locations) and buildings (i.e., building area, building cost, building age) were gathered. In addition, the risk of natural disasters was investigated and linked with the corresponding location information. Table 1 summarizes a set of variables collected. First, capital assets, number of students, and geographic location were grouped as the basic information of the educational facilities. Geographic locations were then classified as rural, urban, or metropolitan. Second, building information includes the area, cost, and age of each educational facility. Additionally, major causes of maintenance and repair were classified as nominal variables. Third, the natural disaster grades of tropical cyclone, tornado, lightning, hail, flood, and storm surge events were applied to evaluate the risk of natural hazards. For the quantitative and objective assessment of natural hazards, the risk rating of the Munich Reinsurance Company's Natural Risk Assessment Network (NATHAN) was adopted. NATHAN was developed through a comprehensive analysis of public sources, scientific analyses, and the severity and frequency of past natural hazards, aimed at estimating the risk of natural hazards around the world through the risk of various natural hazards according to geographical locations [23]. The risk rating of each natural disaster was used as the nominal variable. The descriptive statistics of the variables are shown in Table 2.

Table 1. Descriptions of the variables.

Variable	Description	Data Type
Maintenance and repair cost ratio	Amount ratio used for maintenance and repair	Continuous
Capital asset	Total amount of capital assets (\$)	Continuous
Student	Number of students	Continuous
Location	Location of the facility	Nominal 1: Rural area 2: Urban area 3: Metropolitan area
Building area	Total area of building (acres)	Continuous
Building cost	Total price of the building (\$)	Continuous
Building age	The age of the building at the time the loss occurred	Continuous
Loss cause	The main cause of loss	Nominal 1: Vehicle 2: Theft 3: Vandalism 4: Maintenance 5: Lightening 6: Snow 7: Water 8: Other 9: Wind storm 10: Fire 11: Hail storm
Tropical cyclone	Risk rating of tropical cyclone based on the maximum wind speed in a 100-year return period at the facility	Nominal 0: Zone 0 1: Zone 1 2: Zone 2 3: Zone 3 4: Zone 4 5: Zone 5
Tornado	Risk rating of tornado based on the tornado frequency and intensity interpolation at the facility	Nominal 1: Zone 1 2: Zone 2 3: Zone 3 4: Zone 4
Lightning	Risk rating of lightning based on the lightning frequency per km ² per year at the facility	Nominal 1: Zone 1 2: Zone 2 3: Zone 3 4: Zone 4 5: Zone 5 6: Zone 6
Hail	Risk rating of hail storm based on the hail storm frequency and intensity interpolation at the facility	Nominal 1: Zone 1 2: Zone 2 3: Zone 3 4: Zone 4 5: Zone 5 6: Zone 6
Flood	Flood rating according to FEMA flood area classification	Nominal 0: FEMA flood zone 0 1: FEMA flood zone 100 2: FEMA flood zone 500
Storm surge	Risk rating of storm surge based on the risk of storm surge by return period at the facility	Nominal 0: 100-year return period 1: 500-year return period 2: 1000-year return period

Table 2. Descriptive statistics.

Variables	N	Minimum	Maximum	Mean	Std. Deviation
Maintenance and repair cost ratio	1213	3.37	20.69	12.36	2.70
Capital asset	1213	10.17	20.15	14.30	1.70
Building area	1213	11.36	16.71	15.06	0.84
Building cost	1213	15.26	24.10	21.59	1.11
Building age	1213	2	226	96.62	52.36
Number of students	1213	224	52,300	24,140.67	13,361.62
Location	1213	1	3	2.62	0.73
Loss cause	1213	1	11	6.90	2.08
Tropical cyclone	1213	0	3	0.43	0.99
Tornado	1213	1	4	2.46	0.73
Lightning	1213	1	4	2.18	0.80
Hail	1213	1	3	1.93	0.39
Flood	1213	0	2	0.16	0.48
Storm surge	1213	0	2	0.39	0.79

5. Deep Neural Networks to Learn Maintenance and Repair Costs of Educational Facilities

Deep learning has been widely applied in prediction and recognition studies as one of the machine learning techniques that employ regression or type classification approaches. In general, deep learning algorithms have a neural network that consists of several layers and various structures [24]. A typical learning framework of deep learning models is drawn in the same way as other types of neural networks, but it highlights the use of multiple hidden layers to learn numerous datasets more effectively, in addition to input and output layers [25]. Deep learning algorithms can be mainly classified into five different structure and processing methods: (1) deep neural network (DNN); (2) generative adversarial network (GAN); (3) recurrent neural network (RNN); (4) convolutional neural network (CNN); and (5) autoencoder (AE) [26]. A DNN is a standard neural network with a depth that is determined by the number of hidden layers between the input and output layers [27]. A GAN is underpinned by a generator that produces artificial data to be identical to real data and a discriminator that detects real or false data, which is competitive for image or text-based data systems. Similarly, a CNN is powerful for recognizing, analyzing, and segmenting image data. An RNN is widely used for time-based data prediction systems, and an AE operates automatically based on input data first before it takes activation functions.

Among these, a DNN is designed to identify specific functions of multiple layers and trained to model complex nonlinear relationships [28]. Although a DNN may be susceptible to overfitting, it is largely used for prediction and cataloging in various industries and academic areas [29]. Given the complex nonlinearity of datasets collected in this study, a DNN model for learning the maintenance and repair costs of educational facilities was developed. The learning performance of 14 different DNN alternatives (i.e., 7 different cases of hidden layers \times 2 different dropouts for each hidden layered structure) were then evaluated on the basis of its root mean square error (RMSE) and mean absolute error (MAE). RMSE and MAE are typical indicators that express the magnitude of the error by comparing the prediction result of the artificial neural network model with the real result [30]. The RMSE is underpinned by a quadratic scoring method to measure the average value of error's magnitude, as seen in Equation (2). Comparing it with RMSE, MAE is calculated by weighting all the errors equally and linearly, as shown in Equation (3).

The results of the RMSE and MAE are interpreted as when the error value is closer to zero, the predictability increases.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

where y_i is the i th actual value, \hat{y}_i is the corresponding predicted value, and n represents the number of samples used.

5.1. Pre-Processing for Deep Learning

A total of 1213 data points were collected for 13 input variables (i.e., capital asset, building area, building cost, building age, number of students, location, cause of loss, tropical cyclone, tornado, lightning, hail, flood zone status, and storm surge status). The data were scaled using the z-score normalization method for data preprocessing. Data preprocessing aimed to adjust the range of units and quantities that are difficult to compare. The preprocessed input data was divided into training (learning and validation data) and test sets. The training set is intended to learn datasets deeply, while validating the data is for learning if its performance is optimal. The test set is used to evaluate whether the model is suitable for prediction purposes. It is well known that a 60:20:20 split is generally used for the training, validation, and test set, respectively, within a large set of original data. Given the smaller dataset gathered in this study for deep learning, to improve the efficiency and effectiveness of learning performance, a 70:30 holdout method was adopted by dividing the data into 70% training and 30% test sets. In order to tune hyperparameters, 30% of the training set was assigned to the validation set, which was not used for training.

5.2. Setting Key Components of Deep Neural Networks (DNN) and Exploring DNN Alternatives

The DNN model updates the weights of the neural network nodes and uses a back-propagation algorithm to adjust the model. The optimum arrangement of the model is identified through a trial-and-error basis since the optimal combination is dependent on the input and output variables. In this study, a total of 14 different network structure scenarios were considered for the learning model selection process, in line with two different hyperparameter tuning values for seven different hidden layered network structures.

In a network structure scenario, the number of layers and nodes is determined. The hyperparameters include the optimizer, activation function, dropout, batch size, and epoch [31]. The optimizer is a way to make learning both stable and fast, and the activation function is a way of discovering the weight of each node to arrive at an optimal output (i.e., maintenance and repair cost ratio). The dropout is a way to prevent overfitting with a normalization penalty. If the model is overfitted, its performance deteriorates due to a rise in errors in the test data due to excessive learning of the training data. Overfitting sometimes occurs when the learning model is complicated by a large number of hidden layers and nodes, or when there are many input variables. To identify the overfitting, the original data can be classified into two subsets for training and testing. By using these two data sets, the learning performance is measured by looking at the percentage of accuracy shown in both data sets, which could identify the presence of overfitting. More specifically, if the training performance is better than the data in the test set, it represents that the model tends to be overfitted. A batch is an assemblage of data that is used for efficient calculations. The epoch defines the number of learning periods [31,32]. The network structure scenarios in this study were structured in 3 hidden layers. In general, 1 to 2 hidden layers in deep learning are sufficient if the characteristics of the data are less complex. However, as stated earlier, this study adopted various types of data (i.e., continuous, ordinal, and nominal) for deep learning. Given the complexity of the data, 3 hidden layers were considered for this

study, as a way to achieve the optimal solution. It was assumed that more than three layers increase the possibility of overfitting.

A dropout of 0 or 0.2 was then determined considering the amount of training data, and simulations were conducted to find an optimal combination. In addition, the optimizer adopted the Adaptive Moment Estimation (Adam) method, while the ReLu (Rectified Linear Unit) function was used as an activation function. The ReLu is an activation function that yields 0 if the input value is smaller than 0, and leaves the input value as it is if the input value is larger than 0. This was developed to address the problem of the loss of the slope of the existing Sigmoid function [33]. The Adam Method, an optimization algorithm, is the first-order gradient algorithm that adds the conception of moment to a stochastic objective function. As one of the widely-used algorithms, it was developed in 2015 with regard to computational efficacy and affluence of applications [34]. The batch was set to 5, and the number of epochs was set to 1000. Learning was halted in an epoch when there was no more enhancement of the maintenance and repair cost ratio values.

5.3. Selecting the Final DNN Model

As seen in Table 3, the network structure scenario results clearly present the trend of accuracy associated with MAE and RMSE with two different dropout values of 0 and 0.2. The structures ranging from 25-25-25 hidden layers to 200-200-200 hold larger error values, which is less accurate than the 5-5-5 and 10-10-10 structure scenarios. Hence, 5-5-5 and 10-10-10 were nominated as the finalists of the DNN model. Compared to 10-10-10, the 5-5-5 hidden layered model resulted in producing larger error values. Given the trend of errors between these two structures, it can be expected that any number of hidden layers greater than 5 and less than 10 are less accurate than the final structure of 10-10-10. This investigation concludes that the 10-10-10 hidden-layer DNN structure (a dropout value of 0.2) is the optimal model for this study. Accordingly, the final model is revealed in Table 4, along with the confirmed network structure and the hyperparameters.

Table 3. Learning performance of 14 different deep neural network (DNN) alternatives: Mean absolute error (MAE) and root mean square error (RMSE).

Network Structure Scenario	Dropout (0)		Dropout (0.2)	
	MAE	RMSE	MAE	RMSE
5-5-5	3.630	4.587	2.424	2.846
10-10-10	3.616	4.476	2.371	2.789
25-25-25	5.235	6.536	2.434	2.835
50-50-50	5.303	6.543	2.442	2.840
75-75-75	5.378	6.598	2.447	2.845
100-100-100	5.431	6.602	2.453	2.854
200-200-200	5.975	6.830	2.454	2.852

Table 4. Configuration of the final DNN model.

Group	Composition	Detail
Network Structure	Number of Hidden Layers	3
	Node	10-10-10
Hyperparameter	Optimizer	Adaptive Moment Estimation Method
	Activation Function	Rectified Linear Unit function
	Dropout	0.2
	Batch Size	5
	Epoch	1000

6. Robustness Validation of the Model

Traditionally, a multiple regression analysis (MRA) method has been widely used for quantitative prediction model development. To scientifically validate the robustness of the final DNN model, MRA was conducted further. Then, the prediction results of the MRA were compared to the DNN model by measuring their own MAE and RMSE values and comparing their error values. To improve the quality of the MRA, satisfying the essential assumption of normal distribution, particular variables were log-transformed, including maintenance and repair costs, capital assets, building areas, and building costs. As this study highlights the DNN-driven analytical framework, the MRA process has been omitted.

Table 5 summarizes the results of the robustness validation. First, the overfitting issue on the developed DNN model was tested and validated. More specifically, the validation data results were an MAE of 2.028 and an RMSE of 2.331, while the test data results were an MAE of 2.001 and an RMSE of 2.228. These results indicate that the issue of overfitting is negligible because the difference in the error values between the datasets is not far from each other. When it comes to the robustness validation of the DNN model, it produced lower error rates of 9.1% for MAE and 8.5% for RMSE, compared to the MRA. This concludes that the developed DNN model is more powerful for learning the maintenance and repair costs of educational facilities associated with the multi-contextual factors applied for this study.

Table 5. Model validation results: DNN vs. Multiple regression analysis (MRA).

Model	Validation		Test	
	MAE	RMSE	MAE	RMSE
DNN	2.028	2.331	2.001	2.228
MRA	-	-	2.184	2.417
DNN/MRA (%)			−9.1%	−8.5%

7. Discussion

This study proposed a robust modeling framework by developing a DNN model that learns educational facility maintenance and repair costs from various factors. The developed model was scientifically validated by comparing it with a conventional multiple regression analysis method on the aspects of RMSE and MAE. When comparing the prediction results of the final learning model with those of traditional MRA, the DNN model had a lower prediction error rate than the MRA method. Therefore, to appropriately reveal the nonlinear features of the maintenance and repair costs of educational facilities, the nonparametric DNN model is a better fit than the MRA model, which can be translated into enhanced management of educational facilities by proactively identifying the potential risk of financial losses and enabling active investments in educational facilities. The main finding of this study is that the proposed analytical modeling framework was effective in learning educational facility maintenance and repair costs, which is comparable to other previous studies focused on a limited number of factors affecting maintenance and repair management.

Nevertheless, it should be noted that there are limitations of this study. This study used the claim payout records of one insurance company. Due to the limited dataset available, it was difficult to collect new sets of data experimentally. Further research is needed to compare and verify claim payout records from different insurance companies in the future. In order to more accurately predict maintenance and repair costs using deep learning algorithms, supplementary studies are desired to advance the model by increasing the amount of available data and inputting additional variables, such as maintenance and repair activities and cost components of those activities.

8. Conclusions

Considering the characteristics of current facilities, the need for efficient facility maintenance and management strategies is being emphasized day after day. Effective facility management can have a positive effect not only on the operation and maintenance of facilities, but also on the residents of those facilities. The maintenance of educational facilities is especially vital because proper maintenance and repair of educational facilities can have a great influence on the quality of education. In this sense, it is important to accurately know about maintenance and repair costs.

To address this need, many research efforts have been made by evaluating potential risks within certain facilities and their impacts on maintenance costs. However, the factors applied in previous studies are limited, which are incapable of assessing maintenance and repair costs systematically. For effective facility operations and maintenance management, not only facility replacement/repair and equipment deterioration but also a variety of factors that can decrease the functioning of the actual facility (e.g., accidents, fires, crimes, theft, vandalism, typhoons, earthquakes, floods, hail, heavy snow, and other natural disasters) should be planned for in advance. However, there is still very little known about the impacts of these critical factors that affect maintenance and repairs, although these potential risks are likely to lead to the deterioration or malfunction of some part of a facility or of certain functions. In addition, existing facility operation management research is far from empirical, as most of the calculation of risk for each potential risk factor is done through literature surveys, questionnaires, and expert advice. Therefore, there is still very little known about analytical modeling frameworks that are underpinned by multiple factors affecting facility maintenance and repair costs.

This study attempted to fill these knowledge gaps by modeling and validating a DNN model to efficiently and accurately learn maintenance and repair costs, drawing on 1213 high-confidence data points. The developed model learns and generalizes claim payout records on the maintenance and repair costs in line with facility asset information, geographic profiles, natural hazard records, and other causes of financial losses. By evaluating the learning performance of 14 different DNN alternatives based on their RMSE and MAE values, the 10-10-10 hidden layer with a dropout of 0.2 structure was selected as the optimal learning model. The robustness of the developed model was then scientifically validated by comparing it with a conventional multiple regression analysis method on the aspects of RMSE and MAE. The validation results confirm that the nonparametric DNN model is a better fit than the regression method to appropriately reveal the nonlinear features of the maintenance and repair costs of educational facilities.

This study is the first of its kind and provides a holistic analytical modeling framework that learns high confidence claim payout data from multi-contextual factors affecting maintenance and repair costs of educational facilities. The proposed modeling approach can help industrial practitioners in the discipline of educational facilities consider more critical factors affecting maintenance and repair costs and evaluate their impacts more effectively. The proposed approach can also contribute to the existing body of knowledge, serving as a reference for the facilities management of other functional types of buildings.

Author Contributions: Conceptualization, J.K.; Data curation, S.Y., S.S., and K.S.; Funding acquisition, J.K.; Investigation, J.K. and J.B.; Methodology, J.K. and J.B.; Software, J.K.; Validation, S.Y., S.S., and K.S.; Writing—original draft, J.K. and J.B.; Writing—review and editing, J.B., J.K., S.Y., S.S., and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1F1A1058800).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this research are available from the corresponding or first author by request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lee, K.; Jung, Y. Asset management business function analysis for life cycle information sharing. *J. Korea Constr. Manag. Assoc.* **2016**, *17*, 41.
- Lee, E.; Choi, B.; Jo, Y.; Han, S. Building maintenance prediction model basic plan development using Big Data analysis methodology. *Proc. Korean Inst. Archit. Constr.* **2013**, *13*, 114–115.
- Yoo, Y.; Son, K.; Kim, J.-M.; Kim, T. A probabilistic analysis study on the cost of repair and replacement of educational facilities. *Educ. Facil. Thesis* **2018**, *25*, 3–12.
- Kim, J.-M.; Kim, T.; Yu, Y.J.; Son, K. Development of a maintenance and repair cost estimation model for educational buildings using regression analysis. *J. Asian Archit. Build. Eng.* **2018**, *17*, 307–312.
- Li, P.P.; Locke, J.; Nair, P.; Bunting, A. *Creating 21st Century Learning Environments*; PEB Exchange, Programme on Educational Building; OECD Publishing: Paris, France, 2005.
- Amaratunga, D.; Baldry, D. Assessment of facilities management performance in higher education properties. *Facilities* **2000**, *18*, 293–301. [[CrossRef](#)]
- Woodhouse, J. Optimal timing for replacing aging or obsolete assets. In Proceedings of the Asset Management Conference 2011, IET and IAM, London, UK, 30 November–1 December 2011; pp. 1–10.
- Kim, J.-R.; Jung, Y.-H.; Son, J.-H. A study on reliability analysis model of the repair and replacement cycle of a building which utilizes Monte Carlo simulation. *J. Korea Inst. Build. Constr.* **2010**, *10*, 41–50. [[CrossRef](#)]
- Lavy, S.; Bilbo, D.L. Facilities maintenance management practices in large public schools, Texas. *Facilities* **2009**, *27*, 5–20. [[CrossRef](#)]
- Kaplan, R.S.; Norton, D.P. The balanced scorecard: Measures that drive performance. *Harv. Bus. Rev.* **2005**, *83*, 172.
- Kok, H.B.; Mobach, M.P.; Omta, O.S. The added value of facility management in the educational environment. *J. Facil. Manag.* **2011**, *9*, 249–265. [[CrossRef](#)]
- Tamošaitienė, J.; Peldschus, F.; Al Ghanem, Y. Assessment of facility management candidates by applying game theory. *Procedia Eng.* **2013**, *57*, 1145–1150. [[CrossRef](#)]
- Lavy, S.; Garcia, J.A.; Dixit, M.K. KPIs for facility's performance assessment, Part I: Identification and categorization of core indicators. *Facilities* **2014**, *32*, 256–274. [[CrossRef](#)]
- Yu, K.; Froese, T.; Grobler, F. A development framework for data models for computer-integrated facilities management. *Autom. Constr.* **2000**, *9*, 145–167. [[CrossRef](#)]
- Gilleard, J.D.; Yat-lung, P.W. Benchmarking facility management: Applying analytic hierarchy process. *Facilities* **2004**, *22*, 19–25. [[CrossRef](#)]
- Roper, K.; Payant, R. *The Facility Management Handbook*; AMACOM, HarperCollins Focus: New York, NY, USA, 2014.
- Williams, R.; Shayesteh, H.; Marjanovic-Halburd, L. Utilising Building Information Modelling for Facilities Management. *Int. J. Facil. Manag.* **2014**, *5*, 1.
- Rondeau, E.P.; Brown, R.K.; Lapides, P.D. *Facility Management*; John Wiley & Sons: New York, NY, USA, 2012.
- Gledson, B.J.; Greenwood, D. The adoption of 4D BIM in the UK construction industry: An innovation diffusion approach. *Eng. Construct. Architect. Manag.* **2017**, *24*, 950–967. [[CrossRef](#)]
- IPA. *Transforming Infrastructure Performance, Infrastructure and Projects Authority*; IPA: Hong Kong, China, 2017.
- Kim, S.R.; Jin, R.Z.; Hyun, C.T.; Cho, C.G. Development of Operation & Maintenance Cost Estimating Model for Facility Management of Buildings. *J. Archit. Inst. Korea Struct. Constr.* **2013**, *29*, 11–21.
- Kim, J.-M.; Ha, K.C.; Ahn, S.; Son, S.; Son, K. Quantifying the third-party loss in building construction sites utilizing claims payouts: A case study in South Korea. *Sustainability* **2020**, *12*, 10153.
- Yum, S.G.; Ahn, S.; Bae, J.; Kim, J.-M. Assessing the risk of natural disaster-induced losses to tunnel-construction projects using empirical financial-loss data from South Korea. *Sustainability* **2020**, *12*, 8026.
- Gu, J.Z.; Wang, J.; Kuen, L.; Ma, A.; Shahroudy, B.; Shuai, T.; Liu, X.; Wang, G.; Wang, J.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recogn.* **2018**, *77*, 354–377. [[CrossRef](#)]
- Ajayi, A.; Oyedele, L.; Owolabi, H.; Akinade, O.; Bilal, M.; Davila Delgado, J.M.; Akanbi, L. Deep learning models for health and safety risk prediction in power infrastructure projects. *Risk Anal.* **2020**, *40*, 2019–2039. [[CrossRef](#)]
- Akinosho, T.D.; Oyedele, L.O.; Bilal, M.; Ajayi, A.O.; Delgado, M.D.; Akinade, O.O.; Ahmed, A.A. Deep learning in the construction industry: A review of present status and future innovations. *J. Build. Eng.* **2020**, *32*, 101827. [[CrossRef](#)]
- Nguyen, T.A.; Kashani, T.; Ngo, S.; Bordas, S. Deep neural network with high-order neuron for the prediction of foamed concrete strength. *Comput. Aided Civ. Infrastruct. Eng.* **2019**, *34*, 316–332. [[CrossRef](#)]
- Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
- On Contrastive Divergence Learning. Available online: <https://www.cs.toronto.edu/~hinton/absps/cdmiguel.pdf> (accessed on 1 March 2021).
- Kwon, S.C. Use of frequency analysis of exposure of hazards by occupations: Findings from the third and fourth Korean working conditions survey. *Soonchunhyang Med. Sci.* **2019**, *25*, 37–45. [[CrossRef](#)]
- Bae, S.-W.; Yoo, J.-S. Apartment price estimation using machine learning: Gangnam-gu, Seoul as an example. *Real Estate Stud.* **2018**, *24*, 69–85.
- Ryu, J.-D.; Park, S.-M.; Park, S.-H.; Kwon, C.-W.; Yoon, I.-S. A study on the development of a model for predicting the number of highway traffic accidents using deep learning. *J. Korean ITS Soc.* **2018**, *17*, 14–25.

-
33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. *Imagenet Classification with Deep Convolutional Neural Networks*, NIPS'12; Curran Associates Inc.: Red Hook, NY, USA, 2012.
 34. Kingma, D.P.; Ba, J.L. *ADAM: A Method for Stochastic Optimization*; International Conference for Learning Representations (ICLR): San Diego, CA, USA, 2015.